

HeliCoPTER v1.0:

Heliand Corpus, a Partially Tagged Excel Resource

User Guide

George Walkden

george.walkden@gmail.com

Clare College, Cambridge CB2 1TL

University of Cambridge

October 2011

1. INTRODUCTION	1
2. COLUMNS A-C: THE CLAUSE	2
3. COLUMNS D-G: VERB POSITION.....	3
4. COLUMNS H-K: CLAUSE TYPE	3
5. COLUMNS L-M: MISCELLANEOUS	4
6. COLUMNS N-P: FORCE.....	4
7. COLUMNS Q-X: PROPERTIES OF THE SUBJECT	4
8. COLUMNS Y-Z: MISCELLANEOUS	5
9. KNOWN ISSUES	6

1. Introduction

The HeliCoPTER is an .xls file containing the entire of the *Heliand*, a 5,983-line verse gospel harmony that is our main source for the Old Saxon language. It is separated out by finite clauses (defined as units containing a finite verb), which are then tagged for a number of morphosyntactic properties, of which the most important are:

- Verb position
- Clause type
- Force
- Subject type
- Negation
- Presence of non-finite verbs

The base version is from http://www.hs-augsburg.de/~harsch/germanica/Chronologie/09Jh/Heliand/hel_hf00.html, and is hence based on the Behaghel (1948) edition. As well as being searchable, the HeliCoPTER facilitates finding (and counting) relevant items using Excel's Sort feature.

The HeliCoPTER grew out of my research into verb position in Old Saxon. It is not intended as an end product but will – hopefully sooner rather than later – be made obsolete by the development of a fully syntactically and morphologically annotated corpus of Old Saxon along the lines of the Penn Historical Corpora of English. In the meantime, I am making it available as a stopgap so that other linguists interested in the syntax of Old Saxon can gather data more easily than would otherwise be possible.

The rest of this document outlines the structure of the HeliCoPTER, ending with a section on known issues. If you find examples which in their tagging do not follow the principles laid out here, or other errors, or have any questions, please contact me via email at george.walkden@gmail.com.

2. Columns A-C: The clause

- Column A: N

This is simply a serial number assigned to each clause. Its main purpose is to facilitate sorting the clauses back into order once a different sort has been carried out.

- Column B: Clause

This contains the clause itself. As in the base version, | marks a caesura, and points of divergence between the two main manuscripts are indicated by putting the diverging material within square brackets. Punctuation, of course, should not be assumed to be original. In this version, the finite verb is bold and in italics and / marks a line break. | and / are only retained where they occur clause-medially. ... indicates intervening extraclausal material, as well as marking the beginning and end of clauses not treated by the base version as beginning or ending a sentence. Non-finite clauses are not excluded, but are attached to the most appropriate-seeming finite clause.

- Column C: Lines

This column contains the line on which the clause begins and the line on which it ends, separated by a hyphen if non-identical.

3. Columns D-G: Verb position

These columns are mutually exclusive: all clauses must have one (and only one) of the four. The negator *ni* is treated as a clitic, and adverbial subordinators, complementizers, conjunctions, relative pronouns and *wh*-items preceding indirect questions are not treated as constituents; virtually everything else is treated as a constituent. V1, V2 and V3 are used even when followed by nothing (i.e. verb-final clauses are not necessarily tagged as verb-later).

- Column D: V1
- Column E: V2
- Column F: V3
- Column G: Vlater

4. Columns H-K: Clause type

Values: y or blank.

- Column H: Subordinate
Subordinate and Root are mutually exclusive; all clauses must have one (and only one) of the two. Indirect questions are treated as Subordinate
- Column I: Coord
Any clause may be tagged as Coord in addition to Subordinate (and Relative) or Root. Second and subsequent conjuncts are tagged as Coord; first conjuncts are not. This is because of the well-known observation that second conjuncts in early West Germanic display a high proportion of Vlater orders (see the literature on OE word order). Clauses tagged as Coord may or may not be introduced by a conjunction; clauses introduced by a conjunction should always be tagged as Coord.
- Column J: Relative
All clauses tagged as Relative are also tagged as Subordinate (Relative entails Subordinate, but not vice versa). The two main types are ordinary restrictive relatives and free relatives (SO x SO ...). Apparent free relatives introduced by *Wh*-pronouns (without SO *wh* SO) are always tagged as indirect questions, for simplicity of tagging.
- Column K: Root
It is difficult to distinguish root and subordinate clauses in OS, as in OE. For my purposes, using constituent order as a diagnostic would lead to circularity. I have therefore tried to follow readings given in the base edition.

5. Columns L-M: Miscellaneous

Values: y or blank.

- Column L: Huat
Clauses introduced by ‘interrogative’ HUAT (but not other types of *huat*). There are only 25 of these. A value of y for Huat entails a value of y for Subordinate.
- Column M: Apposition
A stylistic device used extremely liberally in the *Heliand*: two constituents with the same function, which may or may not be linearly adjacent. This column might come in useful for someone at some point.

6. Columns N-P: Force

Values: y or blank.

- Column N: wh-q
A value of y for wh-q combined with a value of y for Root is a direct question. A value of y for wh-q combined with a value of y for Subordinate is an indirect question. The latter category may subsume some free relatives; see under 4.
- Column O: yes-no q
There are only 8 of these disjunctive questions in the whole of the *Heliand*. A value of y for yes-no q entails a value of y for Root.
- Column P: Imperative
A value of y for Imperative entails a value of y for Root.

7. Columns Q-X: Properties of the subject

- Column Q: No subject
Values: y or blank. Used when no nominative constituent eligible for subjecthood is present. A value of y for No subject requires a value of y in one (and only one) of the following: Imperative, *pro*, *con*, Expletive or Gap.
- Column R: Pron
Values: y, (y), ?, or blank. No subject and Pron are mutually exclusive. A value of y indicates that a nominative personal pronoun is the subject of the sentence; other pronominals (such as demonstratives) do not count. A value of (y) indicates that the nominative personal pronoun is present in only one

of the two main manuscripts M and C, as judged by the Sievers (1878) edition. A value of ? indicates that the subject is MAN, which van Bergen has shown to pattern with pronouns in OE (whether it does so in OS has yet to be investigated).

- Column S: Pers
Values: 1, 2, 3. All clauses must have a value. Usually read off the finite verb where the subject itself is not overt.
- Column T: Num
Values: sg, pl. All clauses must have a value. Usually read off the finite verb where the subject itself is not overt.
- Column U: *pro*
Values: y or blank. A value of y is used for clauses with a value of y for No subject that cannot be analysed as involving *con*, Expletive, Gap or Imperative.
- Column V: *con*
Value: y or blank. A value of y is used for clauses with a value of y for No subject and for Coord that cannot be analysed as involving Expletive, Gap or Imperative.
- Column W: Expletive
Value: y or blank. A value of y is used for clauses with a value of y for No subject that cannot be analysed as involving Gap or Imperative and in which the subject is non-referential (in the *Heliand*, mostly clausal subjects).
- Column X: Gap
Value: y or blank. A value of y is used for clauses with a value of y for No subject and for Relative in which it is the subject that has been relativized upon.

8. Columns Y-Z: Miscellaneous

Values: y or blank.

- Column Y: Neg
A value of y is used for clauses in which the preverbal clitic *ni* or *ne* is present.
- Column Z: Multi-verb
A value of y is used for clauses containing at least one non-finite verb in addition to the finite verb.

9. *Known issues*

- Adjacent apposition: When two constituents in apposition are adjacent and preverbal, they may be treated as one constituent or they may be treated as two, with some inconsistency.
- Dative experiencer constructions: In some of these, nominative nominals such as *hugi* and *uulleo* are treated as subjects. In others, a null expletive is assumed to be present. There is inconsistency, but the former strategy is preferred.
- Past participles: Present participles are never treated as non-finite verbs; past participles usually are, but there may be some inconsistency here.
- Distinguishing between subordinate and root: see section 4.
- Subordinate clauses introduced by AND THAT may be tagged inconsistently as coord or not.
- Short unIntroduced clauses in a sequence are tagged as coord when they contain a null subject coreferential with a preceding subject, in order not to overestimate the prevalence of *pro*. In contrast, such clauses are not usually tagged as coord when the subject is overt, and there is no reliable diagnostic for ascertaining whether they are second conjuncts or not.
- Clauses coordinated with wh-q and relative clauses may or may not themselves be tagged as wh-q and relative clauses.
- Overt expletive IT is not distinguished from referential IT.